Structural and energetic study of cation–π–cation interactions in proteins

Silvana Pinheiro,[1] Ignacio Soteras,[2] Josep Lluis Gelpí,[3] François Dehez,[4] Christophe Chipot,[4,5] F. Javier Luque,[2] and Carles Curutchet[1,*]

[1]*Departament de Farmàcia i Tecnologia Farmacèutica i Fisicoquímica and Institut de Biomedicina (IBUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Barcelona Spain*

[2]*Departament de Nutrició, Ciències de l'Alimentació i Gastronomia and Institut de Biomedicina (IBUB), Facultat de Farmàcia i Ciències de l'Alimentació, Universitat de Barcelona, Santa Coloma de Gramenet, Spain.*

[3]*Departament de Bioquímica i Biomedicina Molecular, Facultat de Biologia, Universitat de Barcelona, Spain*

[4]*Laboratoire International Associé Centre National de la Recherche Scientifique et University of Illinois at Urbana −Champaign, Unité Mixte de Recherche No. 7565, Université de Lorraine, Vandoeuvre-lès-Nancy cedex, France*

[5]*Beckman Institute for Advanced Science and Engineering, University of Illinois at Urbana — Champaign, Urbana, Illinois (USA).*

[*]Corresponding author: CC (carles.curutchet@ub.edu)

Abstract

Cation–π interactions of aromatic rings and positively charged groups are among the most important interactions in structural biology. Although the role and energetic characteristics of these interactions is well established, the occurrence of cation–π–cation interactions is an unexpected motif, which raises intriguing questions about its functional role in proteins. We present a statistical analysis of the occurrence, composition and geometrical preferences of cation–π–cation interactions identified in a set of non-redundant protein structures taken from the Protein Data Bank. Our results demonstrate that this structural motif is observed at a small, albeit non-negligible frequency in proteins, and suggests a preference to establish cation–π–cation motifs with Trp, followed by Tyr and Phe. Furthermore, we have found that cation–π–cation interactions tend to be highly conserved, which supports their structural or functional role. Finally, we have performed an energetic analysis of a representative subset of cation–π–cation complexes combining quantum-chemical and continuum solvation calculations. Our results point out that the protein environment can strongly screen the cation–cation repulsion, leading to an attractive interaction in 64% of the complexes analyzed. Together with the high degree of conservation observed, these results suggest a potential stabilizing role in the protein fold, as demonstrated recently for a miniature protein (Craven *et al*. *J. Am. Chem. Soc.* **2016**, *138*, 1543). From a computational point of view, the significant contribution of non-additive three-body terms challenges the suitability of standard additive force fields for describing cation–π–cation motifs in molecular simulations.

Keywords*:* Cation–π–cation complexes; noncovalent interactions; cooperativity; protein structure.

1. INTRODUCTION

The comprehension of intermolecular forces is key to understand the behavior of chemical systems at the molecular level. This is well exemplified in biological systems, given that the molecules of life (DNA, RNA and proteins) are held in their three-dimensional structures by a variety of intermolecular noncovalent interactions, and that mutational changes in proteins can affect the structure and dynamics, and even ultimately the evolution toward novel functions in enzymes.[1] Because the three-dimensional molecular structure of these molecules is responsible for their specific biological activity, understanding noncovalent interactions is fundamental to rationalize the relationships between structure, dynamics, and function. Furthermore, such knowledge is essential for engineering proteins toward the design of enzymes with novel catalytic roles.[2,3]

In the last decades, there has been a continuous effort aimed at identifying and characterizing the many types of noncovalent interactions observed in chemical systems, including, for example, non-conventional hydrogen bonds, π–stacking, cation–π and anion–π contacts, as well as halogen bonding.[4–19] Cation–π interactions have emerged as one of the most important interactions in determining specific recognition motifs in biomolecules.[4–7,11,12] The term cation–π describes the interaction between a positively charged atom or group and a π–electron system. Theoretical studies have elucidated the nature of cation–π interactions, establishing that the leading term involves the electrostatic attraction between a monopole (the cation charge) and the quadrupole created by the π–electron cloud of the aromatic ring.[6,7,13,20–22] The presence of a positive charge near a highly polarizable π–system, however, leads also to important induction contributions to the interaction energy.[6,7,13,20] In addition, dispersion contributions, although smaller, can also be important depending on the nature of the cation involved in the interaction.[13,20] On the other hand, solvation and more generally environment effects may have a crucial influence on the stability of cation–π complexes given the charged nature of the cation.[23] Cation–π interactions are not properly described in standard pairwise additive force fields used in biomolecular simulations, where polarization effects are only taken into account in an implicit way. This shortcoming can be naturally overcome by switching to an explicit polarizable force field,[24,25] though some research groups have also proposed empirical corrections

to additive force fields to improve the description of cation–π interactions.[7,26–30] Such corrections, nevertheless, cannot account for nonadditive effects, which are important to describe such interactions.[31]

There is a large body of evidence illustrating the common occurrence of cation–π interactions in biological systems, including peptides, nucleic acids, enzymes, antibody-binding interactions or ligand-protein binding.[12] Statistical analysis of the structures deposited in the Protein Data Bank (PDB)[32] has unveiled the occurrence and geometric characteristics of cation–π interactions in biomolecules.[7] In addition, some research groups have shown the existence of cooperative effects among cation–π interactions and hydrogen bonds or π–stacking.[9,33–42] Despite all these advances in characterizing cation–π motifs and their interplay with other interactions, a special case of cation–π interaction involving three molecular species, termed cation–π–cation motif, has hitherto received little attention and is still poorly understood. The ability of cation–π–cation interactions to stabilize tertiary structures of proteins has been recently demonstrated for a miniature protein, which emulates a network of cation–π interactions found in fibronectin type III domain containing proteins.[43]

In this contribution, we aim at advancing our understanding of cation–π–cation interactions and their role in the structure and stability of biomolecular systems via two complementary approaches. First, a thorough statistical analysis of the frequency, composition and geometrical features of cation–π–cation complexes identified in a set of non-redundant protein structures from the PDB is presented. Specifically, interactions between the π–electron system of the aromatic residues with not only cationic amino acids, but also different monovalent and divalent cations have been considered. We have also analyzed the degree of conservation of the interactions based on the probability of residue replacement as well as the relations among the protein and its homologs. Second, the most relevant contributions to the interaction energy in cation–π–cation interactions from high-level quantum mechanical (QM) calculations have been analyzed. This study demonstrates that cation–π–cation interactions are unexpectedly quite common in proteins, as about 7% of the structures investigated have one or more cation–π–cation interactions. The results also indicate that the stability of a cation–π–cation complex, which is intrinsically repulsive, is largely dependent on the local environment. Finally, non-additive three-body contributions are significant in cation–π–

cation interactions, implying that the proper description of these motifs in standard additive pairwise force fields is even more challenging from a theoretical perspective compared to cation–π pairs.

The present contribution is organized as follows. First, we discuss the methodological approach used in this work and the computational details of the QM calculations. Second, we report the results of the statistical analysis performed on the set of non-redundant proteins of the PDB, as well as the analysis of the conservation of the most relevant cation–π–cation motifs. Then, we discuss the energetic analysis of the most representative subset of these interactions. Finally, we close with future prospects and possible extensions of the work.

2. METHODS

2.1 Statistical analysis

The statistical analysis of cation–π–cation interactions was performed over a non-redundant protein structure dataset consisting of *ca.* 21000 protein structures taken from the PDB's Cluster90.[32,44] We performed an exhaustive search for cation–π–cation motifs established between the side chains of aromatic residues (Phe, Tyr, Trp) with cationic amino acids (Lys and Arg), as well as with different monovalent and divalent cations (Na^+, Ca^{2+}, Mg^{2+}, K^+, Li^+, Cu^{2+}). The criteria to detect a cation–π–cation complex were based on the distance from the cation to the aromatic ring, as well as on the angle formed by the two cations and the aromatic ring. The latter is defined as the angle between the normal of the aromatic ring and the vector connecting the centroid of the ring and the positively charged site (Figure 1). In the case of Lys and Arg, the distance was taken to the N_ζ atom of the protonated amine and to the guanidine C_ζ atom, respectively. In the case of Trp, both six- and five-membered aromatic rings were considered. The set of interactions initially identified was then filtered in order to remove possible duplicates in each interaction arising from symmetry considerations within the PDB.

For the sake of completeness, we have also extended the analysis to complexes that include His as a potential partner. Whereas most Lys and Arg residues are expected to be cationic, it is well known that His residues generally populate the neutral state,

even though they may be protonated depending on their local environment in proteins. Due to the lack of information regarding the protonation states in the PDB structures, and the range of uncertainty in predicting the pKa of His residues,[45,46] we have decided to extend our statistical analysis to cation–π–His and His–π–His complexes, since even a neutral His might a priori be an unexpected motif to assist in the formation and stability of the complex through hydrogen bonding to the aromatic ring. The protonation state of the His residues participating in the complexes identified was then investigated using the PROPKA server.[45,46] In the case of complexes involving His, the geometrical parameters (Figure 1) were defined to determine the relative position of the nitrogen atoms of the His residue relative to the aromatic ring.

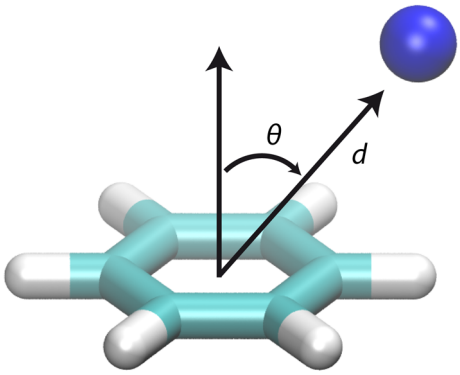

Figure 1. Definition of the geometrical parameters that characterize the cation–π interaction between a π ring and a positively charged group.

Taking into account the preceding definitions for the center of the aromatic rings and the location of the charged sites, we considered a threshold value for the maximum distance (d) between the cation and the aromatic ring of 5 Å, and an angle cutoff (θ) for the cation–aromatic ring–cation arrangement of 45 degrees (Figure 1). The initial set of interactions was then refined to identify the most relevant ones by imposing a more stringent criterion, consisting of a combined distance and angle cutoffs for both cations, such that $d_1 + d_2 < 8$ Å and $\theta_1 + \theta_2 < 40°$, respectively. From this new subset, we removed (i) complexes containing His residues, due to the uncertainty related to its protonation state, (ii) Lys complexes where the ammonium group was largely displaced from the normal axis to the aromatic ring, and (iii) proteins with a high sequence

similarity with other proteins in the set. Some proteins of this set featured multiple interactions (PDB IDs 1GXS, 1T72, 2H39, 2J47, 2JKV, 2PAM and 3G9V), which were kept, although they did not meet our geometric criteria, as we deemed interesting to investigate the energetic characteristics of these special cases. Moreover, we also included in the final set of interactions a particularly interesting system, namely the human growth hormone receptor (PDB entry 1A22), where three consecutive cation–π–cation interactions are established involving amino acids Lys379, Trp386, Arg411, Phe425, Arg413, Tyr422, and Lys415, as shown in Figure 2. The final refined dataset included 53 interactions (see Results section). For this final set of interactions, we analyzed the degree of conservation of the amino acids involved in the complex, and performed the energetic analysis described in the following section.

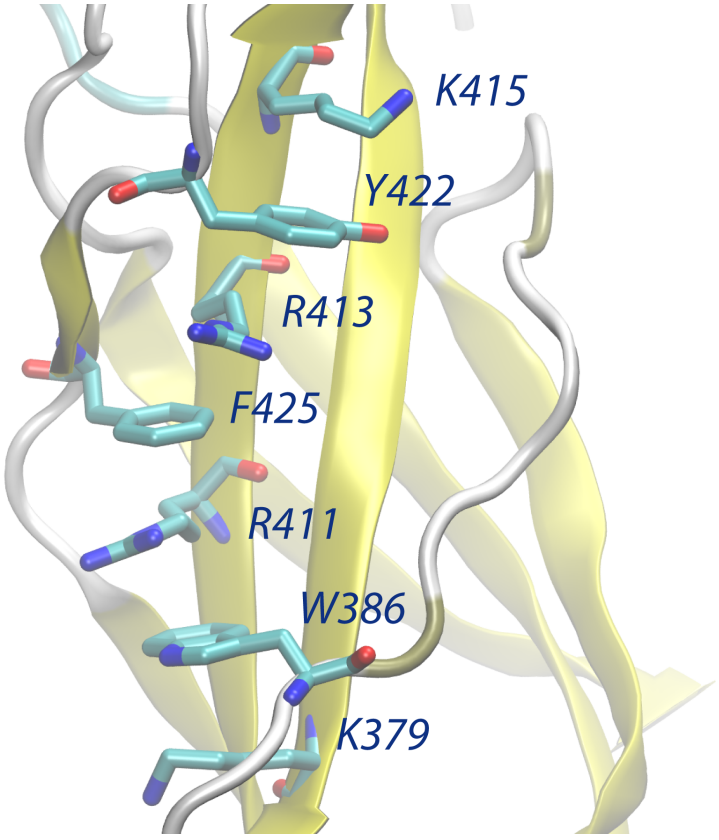

Figure 2: Multiple cation–π–cation interactions identified in the human growth hormone receptor (PDB entry 1A22).

The degree of conservation of the cation–π–cation interactions in each protein was determined with the aid of the ConSurf web server,[47,48] which accounts for the

probability of residue replacement, as well as the evolutionary relations among the protein and its homologs. Finally, we also computed the pKa and the fraction of buried surface area of the titratable amino acids in the cation–π–cation interactions using the PROPKA server.[45,46] In all cases, the Arg and Lys amino acids were found to be protonated at neutral pH (see Table S7 in the Supporting Information).

2.2 Energetic analysis

For each complex featured in the subset of cation–π–cation motifs, a simplified model was built up to perform the energetic analysis using high-level QM calculations. To this end, Lys and Arg amino acids were represented as ammonium and guanidinium ions, whereas Phe, Tyr, and Trp were represented as benzene, phenol, and indole units, respectively. The geometries of these prototypical molecules were fully optimized at the MP2/aug-cc-pVDZ level using the Gaussian 09 suite of programs.[49] The optimized geometries were then aligned onto the corresponding fragments of the amino-acid side chains, thereby forming the cation–π–cation motif as found in the original PDB structures. For the models involving Lys, the orientation of the hydrogen atoms in the ammonium group was then optimized through MP2/aug-cc-pVDZ calculations, while keeping the rest of the atoms frozen in the model system.

The trimer interaction energy, $\Delta E_{XArY} = E_{XArY} - E_X - E_{Ar} - E_Y$, as well as the corresponding pair interaction energies, $\Delta E_{XAr} = E_{XAr} - E_X - E_{Ar}$, $\Delta E_{ArY} = E_{ArY} - E_{Ar} - E_Y$ and $\Delta E_{XY} = E_{XY} - E_X - E_Y$, were determined to assess the contribution of three-body non-additive effects to the interaction according to the following expression:

$$\Delta E_{\text{3-body}} = \Delta E_{XArY} - \Delta E_{XAr} - \Delta E_{ArY} - \Delta E_{XY} \quad (1)$$

where Ar is the aromatic amino acid and X,Y denote the cations involved in the X–Ar–Y cation–π–cation interaction.

Single-point energies for monomers, dimers and trimers were calculated using the spin-component-scaled Møller–Plesset second order (SCS-MP2) method proposed by Grimme, which has been shown to improve significantly the accuracy of MP2 calculations.[50] Essentially, the SCS modification involves scaling parallel and anti-

parallel-spin components of the MP2 correlation energy according to the following expression:

$$E_{SCS-MP2} = E_{HF} + \frac{1}{3}(E_{corr(\alpha-\alpha)} + E_{corr(\beta-\beta)}) + \frac{6}{5}E_{corr(\alpha-\beta)} \quad (2)$$

In addition, all SCS-MP2 energies were extrapolated to the complete basis set limit (CBS) combining results obtained using Dunning's cc-pVDZ and cc-pVTZ basis sets according to the formula proposed by Truhlar:[51]

$$E_{SCS-MP2/CBS} = \frac{3^{\alpha}}{3^{\alpha}-2^{\alpha}}E_3^{HF} - \frac{2^{\alpha}}{3^{\alpha}-2^{\alpha}}E_2^{HF} + \frac{3^{\beta}}{3^{\beta}-2^{\beta}}E_3^{corr} - \frac{2^{\beta}}{3^{\beta}-2^{\beta}}E_2^{corr} \quad (3)$$

where E_2 and E_3 denote Hartree-Fock and correlation energies obtained using double-ζ and triple-ζ basis sets, and $\alpha = 3.4$ and $\beta = 2.2$.

For the 3 cation–π–cation interactions identified in the human growth hormone receptor (PDB entry 1A22), interaction energies were also computed at the coupled cluster with single and double excitations (CCSD) level of theory using the aug-cc-pVDZ basis set. In all cases, dimer and trimer interaction energies were corrected for basis-set superposition errors (BSSE) using the counterpoise correction (CP),[52] in which monomer calculations were performed using the complete basis set of the interacting dimer or trimer systems.

Finally, the effect of the environment on the cation–π–cation complex was examined using two strategies. First, preliminary insight into the effect of the environment was gained from QM continuum solvation calculations in water and in *n*-octanol using the MST-IEFPCM B3LYP/6-31G(d) model, following the scheme described in previous studies.[53–55] Next, we estimated the contribution of the protein environment by computing the free-energy change associated to transferring the trimer from water to the protein scaffold using the classical MM-GBSA method, as implemented in the Amber12 software.[56] In particular, we used the modified GB model developed by Onufriev, Bashford and Case using a dielectric constant of 78.3.[57] In order to avoid steric clashes between the protein and the model trimer in the latter calculations, Arg and Lys residues participating in the interaction were mutated into Gly, and the aromatic residue (Trp, Phe or Tyr) was removed, capping the neighboring amino acids with a hydrogen atom. In addition, all cofactors were removed from the

structures, and the protonation states of the residues in the polypeptide chain were assigned according to PROPKA calculations.[45,46] In the particular case of the phosphate transport phoU protein (PDB entry 1T72), the low resolution of the crystal structure led to artificially large van der Waals contributions in the MM-GBSA calculations of its two cation–π–cation complexes. Therefore, before performing the MM-GBSA calculations, we minimized the protein in vacuum keeping the amino acids involved in the interactions restrained to their initial position. All MM calculations were based on the parm99SB Amber force field,[58] and the parameters for the ammonium, guanidinium, benzene, phenol and indole units were assigned using the Antechamber module based on the Amber force field with RESP charges derived at the HF/6-31G(d) level of theory.[59]

3. RESULTS AND DISCUSSION

3.1 Statistical analysis of cation–π–cation motifs in proteins

We performed an initial analysis of the non-redundant protein dataset in search for cation–π–cation, cation–π–His and His–π–His motifs. Interactions were identified when the distance and angle formed by the cationic moeities relative to the aromatic ring were below 5 Å and 45°, respectively (Figure 1). These criteria led to the identification of a total of 2,898 interactions located in 2,328 protein structures, which correspond to 1,675 cation–π–cation, 1,029 cation–π–His and 194 His–π–His motifs. Investigation of the protonation state of the His residues involved in the cation–π–cation motifs suggested a cationic His participating in 24 cation–π–His^+ and 10 His^+–π–His motifs, whereas no His^+–π–His^+ complexes were found. Thus, our results indicate an overall number of 1699 cation–π–cation complexes found in 1,450 proteins (7.0% of the structures analyzed). This result stresses the unexpectedly significant population of these interactions in proteins.

Figure 3 shows the statistical distribution of cation–π and His–π distances and angles for the 2,898 complexes. The aromatic ring corresponds to Tyr in 47% of the cases, whereas a similar population is found for Trp (25%) and Phe (28%). Additional detail about the particular composition of each interaction identified is supplied in Tables S1-S3 of the Supporting Information. Moreover, Table 1 shows the minimum and mean values of the distance (d) and angle (θ) between the cation and the aromatic ring. The

distributions of average distances and angles are very similar for the three aromatic amino acids, with a mean distance ranging from 4.1 to 4.4 Å, and an average angle from 27 to 30 degrees.

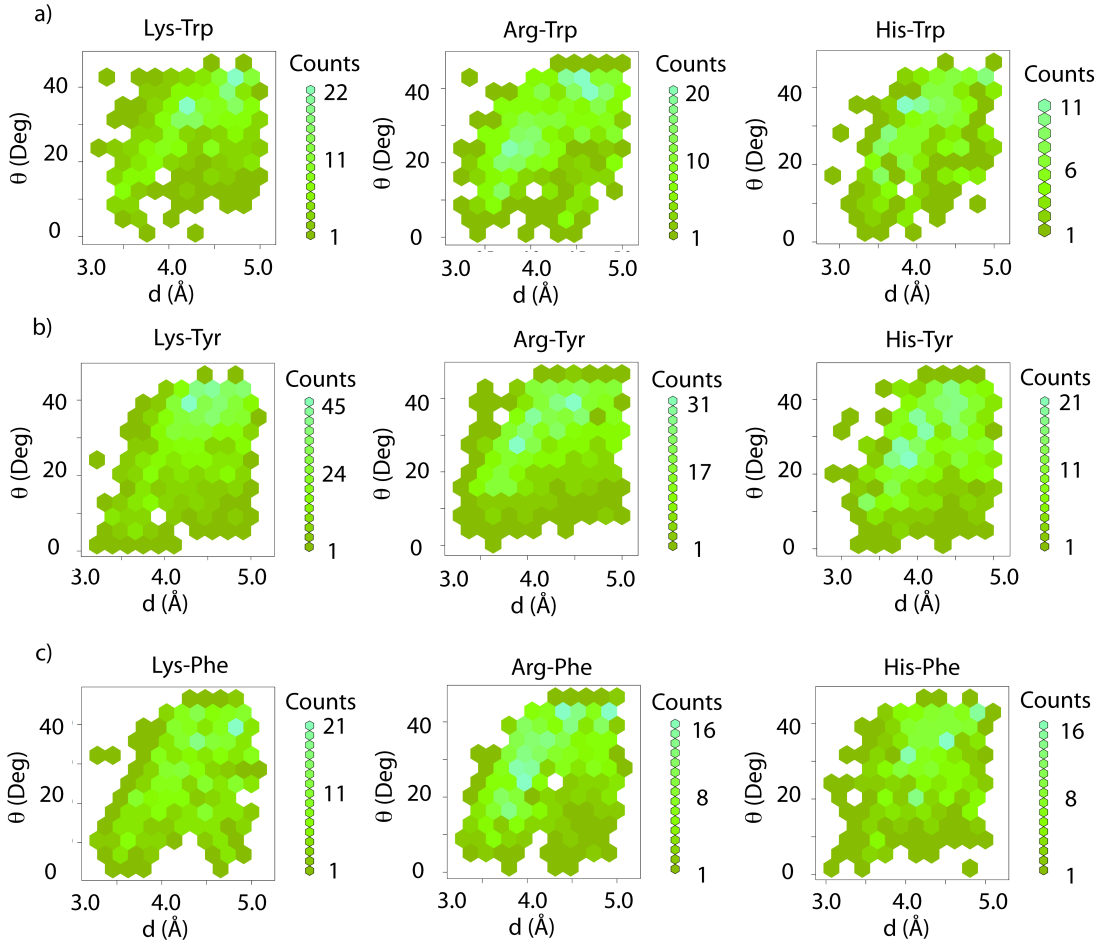

Figure 3. Statistical distribution of pairwise cation(Lys, Arg)–π and His–π distances and angles of approach observed for the 2,898 interactions identified in the Cluster90 database involving a) Trp, b) Tyr and c) Phe.

Table 1. Minimum and mean pairwise cation(Lys, Arg)–π and His–π distances and angles of approach observed for the 2,898 interactions identified in the PDB.

		Lys		Arg		His	
		d (Å)	θ (°)	*d* (Å)	θ (°)	*d* (Å)	θ (°)
Trp	Min	3.23	0.9	3.24	1.0	2.91	2.5
	Mean	4.29	29.2	4.21	27.8	4.11	27.4
Tyr	Min	3.22	1.6	3.29	0.4	2.78	1.2
	Mean	4.35	29.7	4.22	29.1	4.16	27.3
Phe	Min	3.28	3.4	3.28	1.4	3.07	1.7
	Mean	4.33	27.5	4.22	28.0	4.25	28.4

The analysis of the complexes formed by Tyr as the aromatic residue, which is the most frequent cation–π–cation motif within the dataset, primarily involves the mixed interaction with the ammonium group of Lys and the guanidinium moiety of Arg, the Lys–Tyr–Arg motif being detected in 26% of the X–Tyr–Y complexes. Complexes involving the same charged amino acid residue (X–Tyr–X) were found in 38% of the cases, the most favored one being Lys–Tyr–Lys (19%). In very few instances the cation–π–cation interaction involved a metal cation (Ca^{2+}, Na^{+} or Mg^{2+}), generally forming Lys–Tyr–Mg^{2+} complexes. On the other hand, no interaction motifs were found involving the simultaneous presence of two metal cations.

With regard to Trp, the preferred motif with charged amino acids was also the mixed complex formed with Lys and Arg (28% of the cases), thus mimicking the preferences observed for Tyr. Similar trends were observed for Phe, as the preferred motif (23%) is the Lys–Phe–Arg complex. Moreover, concerning cases in which the same amino acid is found on each side of the aromatic ring, the interaction motifs Arg–Trp–Arg (20%) and Arg–Phe–Arg (16%) were found to be the most common, in contrast with the Tyr case. Again, only a few complexes were detected involving the metal cations Na^{+} and Mg^{2+} in the case of Trp, and Ca^{2+} in the case of Phe. As has been noted for Tyr, the simultaneous interaction of both Trp and Phe with two metal cations was not observed in the present analysis.

Overall, we can see that all the considered aromatic amino acids (Trp, Tyr, Phe) show a clear preference to interact with charged amino acids in cation–π–cation interactions rather than monovalent metal cations. This trend can presumably be attributed to the larger hydration free energies of the metal cations compared to the ammonium and guanidinium groups,[60–63] and the larger volume accessibility of the metal cations compared to the pre-organized arrangement of the charged residues in the protein fold, but also to the limited number of protein structures involving metal cations in the dataset. If we discard those complexes involving a neutral His, the preferred amino acid in such interactions is found to be Tyr (48%), followed by Trp (26%) and Phe (25%). Despite its two rings, the fact that Trp appears in a small number of cation–π–cation complexes can be explained by the smaller natural occurrence of Trp (1.3%) compared to Tyr (3.3%) or Phe (3.9%).[64] Indeed, if the number of interactions found is weighted by their natural abundance, our results suggest a preference to establish

cation–π–cation motifs with Trp, followed by Tyr and Phe (approximate ratios 3:2:1). If we consider charged amino acids on each side of the aromatic ring, the dominant motif contains Lys and Arg, and these residues appear with a similar frequency among all the possible motifs. Again, this finding can be explained by the similar abundance of Lys (5.8%) and Arg (5.2%). The smaller number of complexes involving His probably stems not only from its lower natural abundance (2.2%), but also from the intrinsic preference for its neutral form (see above).

Given the large amount of cation–π–cation complexes found in the dataset, we performed a selection procedure in order to identify the most relevant interactions to be used in the energetic analysis. To this end, a more stringent geometrical analysis was carried out looking at the distribution of distances and angles found in the trimeric (X–π–Y) complexes, shown in Figures 4-6.

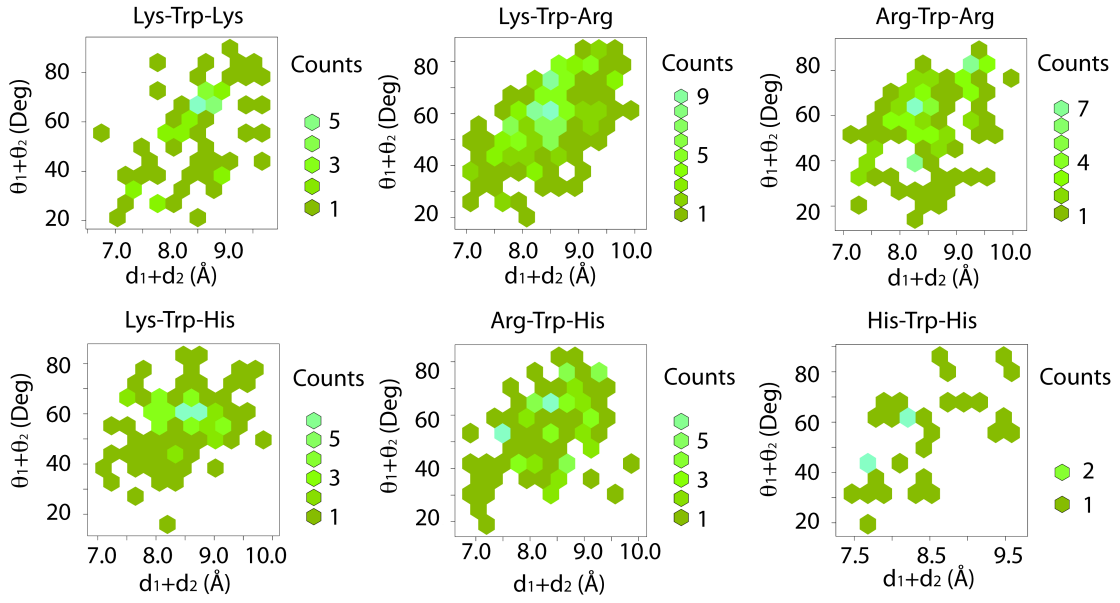

Figure 4. Statistical distribution of distances and angles of approach observed for the trimeric (X–π–Y) complexes identified involving Trp as the aromatic residue.

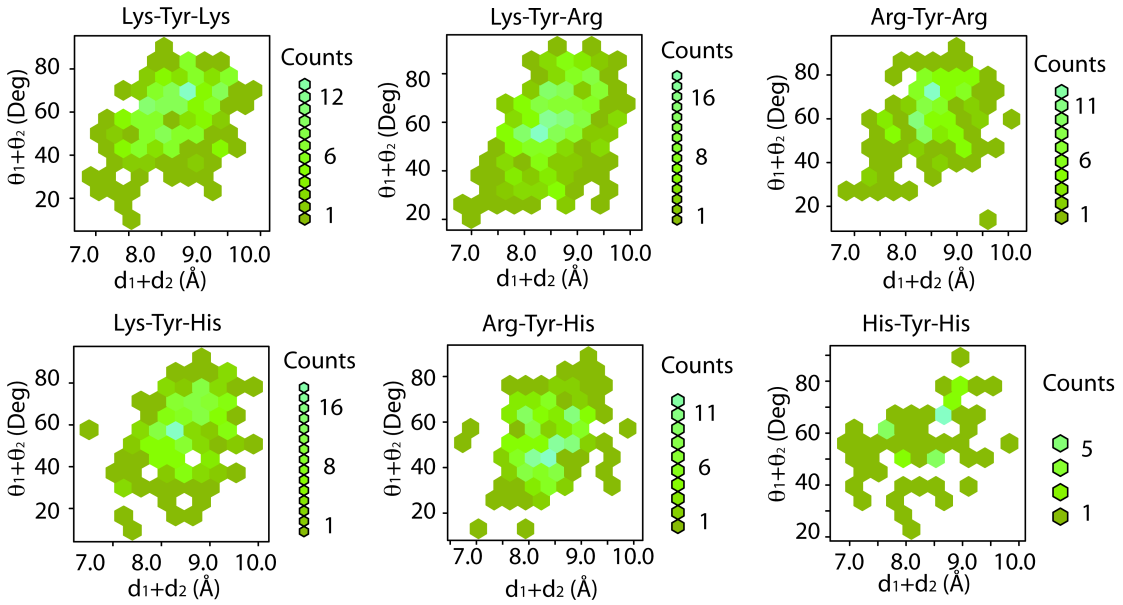

Figure 5. Statistical distribution of distances and angles of approach observed for the trimeric (X–π–Y) complexes identified involving Tyr as the aromatic residue.

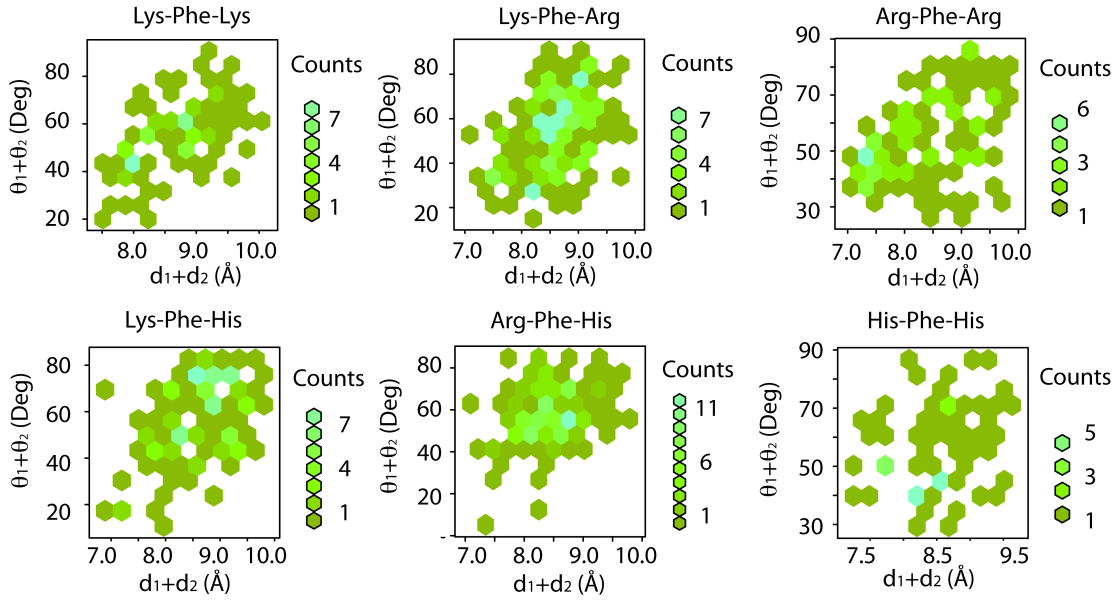

Figure 6. Statistical distribution of distances and angles of approach observed for the trimeric (X–π–Y) complexes identified involving Phe as the aromatic residue.

As can be noted in Figures 4-6, most of the complexes are clustered in the upper right corners, corresponding to total distances ($d_1 + d_2$) and angles ($\theta_1 + \theta_2$) in the range of 9.0-9.5 Å and 60-80 degrees, respectively. However, there are cases with total distances in the range of 7.0-8.0 Å, and angles lower than 40 degrees. For our purposes

here, we filtered out complexes using such criteria (see Methods). These additional criteria led to a selection of 175 complexes. As indicated in the Methods section, this subset was further refined by excluding (i) complexes with His residues, (ii) Lys complexes where the ammonium group was largely displaced from the normal axis to the aromatic ring, and (iii) highly similar proteins. In addition, we kept in our subset specific complexes taken from proteins that featured multiple cation–π–cation interactions, although they did not strictly meet our geometric criteria.

These refinements led to a final selection of 53 complexes (Tables S4-S6 of the Supporting Information). Then, we performed a multiple structural alignment based on the central aromatic residue in order to investigate the three-dimensional distribution of charged residues around the aromatic ring, which is shown in Figure 7. This analysis reveals three trends: (i) the nitrogen atom of Lys does not tend to be faced towards the center of the ring, an effect that may reflect a balance between the cation–π–cation and other interactions formed by the protonated amine with neighboring residues, (ii) the distribution of Arg (as measured from the position of the guanidinium C atom) is closer to the normal to the centroid of the aromatic ring, which reflects both the larger delocalization of the positive charge in the guanidinium moiety and the enhanced contribution of dispersion forces between the guanidinium group and the aromatic ring,[13,20] and finally (iii) in the case of Trp, the preferential ring found in cation–π–cation interactions seems to be the five-membered one, likely reflecting the preference for the negative charge borne by the nitrogen atom.

We also examined the relative orientation of the Arg guanidinium moiety and the aromatic ring. Figure 8 shows the distribution of the angles formed by the molecular planes of these two moieties. As expected, our results show that both planes are roughly parallel, with peaks centered in the range 0-20 degrees for the three aromatic residues. However, in two specific cases involving Trp and Phe complexes, respectively, the guanidinium moiety and the aromatic ring adopted a T-shape orientation (PDB entries 1CN4 and 1GQK).

Finally, we also investigated the degree of conservation of the amino acids found in the set of 53 cation–π–cation interactions with the aid of the ConSurf web server,[47,48] which accounts for the probability of residue replacement as well as the

evolutionary relations among the protein and its homologs. Figure 9 shows the distribution of conservation scores for the set of 53 complexes, obtained as an average over the individual values of each amino acid. Such individual scores, as well as the corresponding residue variety (an index that defines the degree of mutational diversity in ConSurf), can be found in Table S7 of the Supporting Information. The scoring scale ranges from 1 to 9, the latter value indicating the highest degree of conservation. Interestingly, the interactions involving Trp and Phe seem to be strongly conserved, with average scores between 6-9 in most cases. On the other hand, interactions involving Tyr tend to be significantly less conserved.

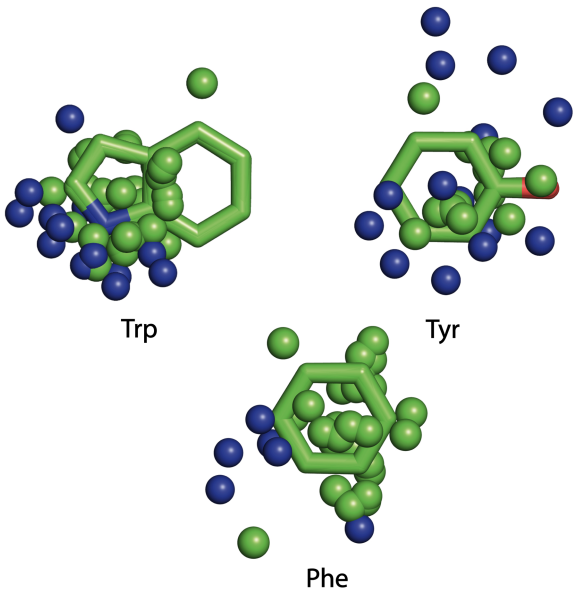

Figure 7. Spatial distribution of Lys and Arg cations around Trp (upper left), Tyr (upper right) and Phe (bottom center). The C_ζ and the N_ζ atoms of Arg and Lys are represented by green and blue spheres, respectively.

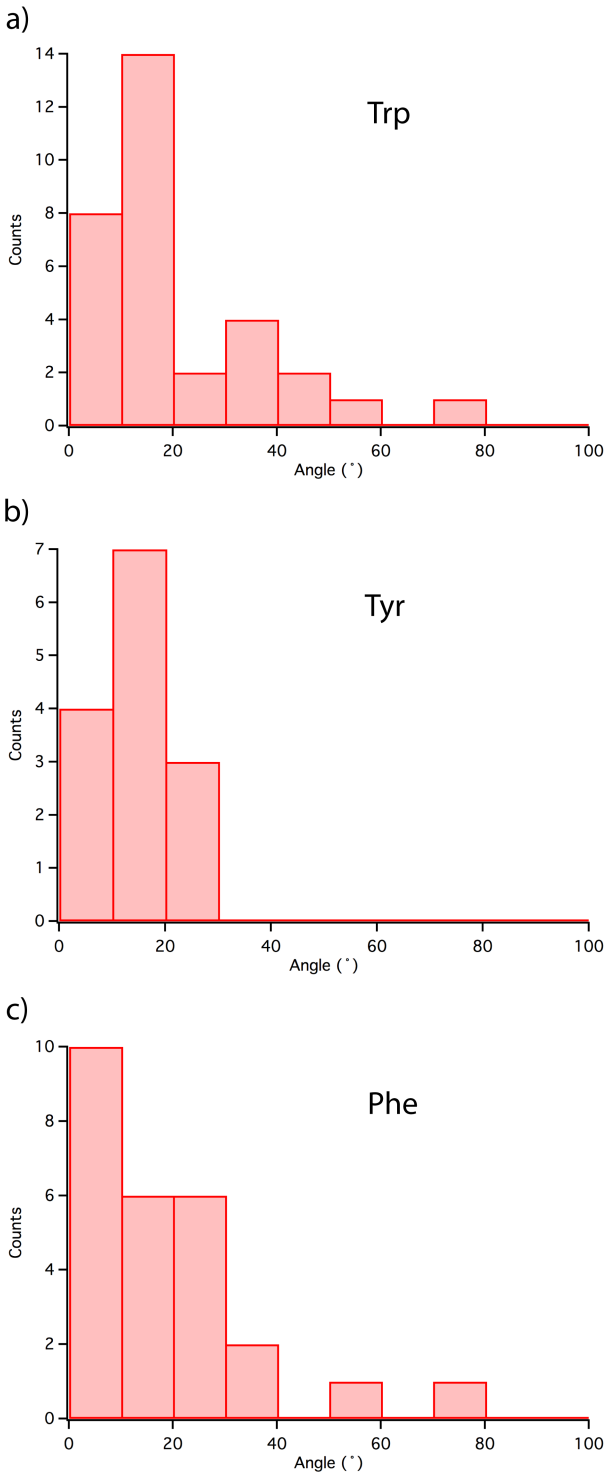

Figure 8. Histograms of the relative angle between the arginine and aromatic planes for cation–π–cation interactions involving a) Trp, b) Tyr and c) Phe.

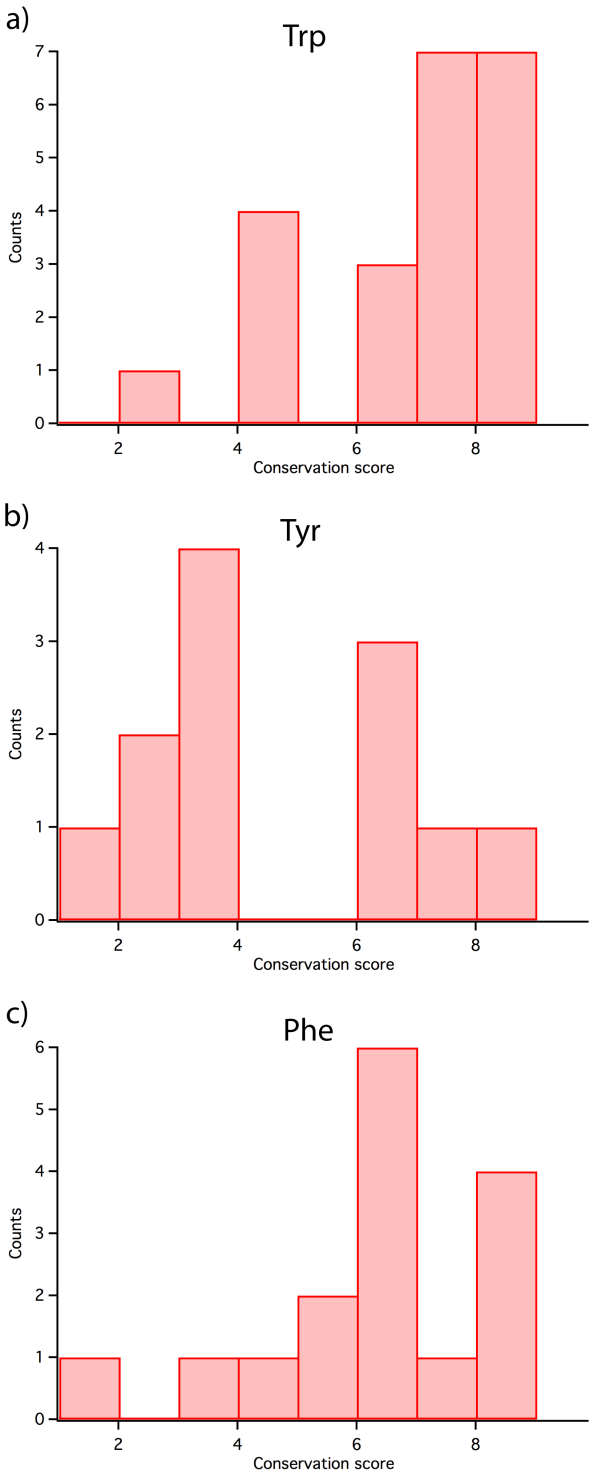

Figure 9. Histograms of the conservation score averaged over the three amino acids for cation–π–cation interactions involving a) Trp, b) Tyr and c) Phe. The scoring scale ranges from 1 to 9, the latter value indicating the highest degree of conservation.

3.2 Energetic analysis

The common occurrence of cation–π–cation motifs identified in our statistical analysis warrants investigating the underlying energetics of such interactions. A particularly challenging question is whether or not they can be expected to stabilize the tertiary structure of a protein. On the other hand, it is of interest to investigate the cooperativity among simultaneous two-body cation–π interactions, which we estimate here by computing three-body non-additive effects in the interaction energies, effects that are envisioned to be nonnegligible, given the well-known impact of polarization effects on cation–π interactions.[6,7,13,20]

Figure 10 shows the distribution of pair interaction energies obtained for the complexes containing Trp, Tyr and Phe at the SCS-MP2/CBS level of theory in the gas phase. In addition, we show the overall distribution of pair interaction energies, as well as the trimer interaction energies and the corresponding three-body contributions (data available in Table S8 of the Supporting Information).

As expected, our results indicate that the interaction energies involving two cations are strongly repulsive, in the range ~30-50 kcal/mol, with similar energetic distributions observed for Lys–Lys, Arg–Arg and Lys–Arg pairs in Trp, Tyr or Phe complexes. On the other hand, the distribution of cation–π interaction energies shows larger differences depending on the specific nature of the aromatic ring involved in the complex. A notable difference is that the distribution of Lys–Phe and Arg–Phe interaction energies is considerably narrower than their counterparts involving Trp or Tyr, probably due to the greater symmetry of Phe. Indeed, the distributions involving Trp, where the cations can interact with both the five- and six-membered rings, exhibit the broadest distributions. Interestingly, regardless of the aromatic amino acid involved, the strongest interaction energies in the distribution always arise from cation–π pairs involving Lys rather than Arg.

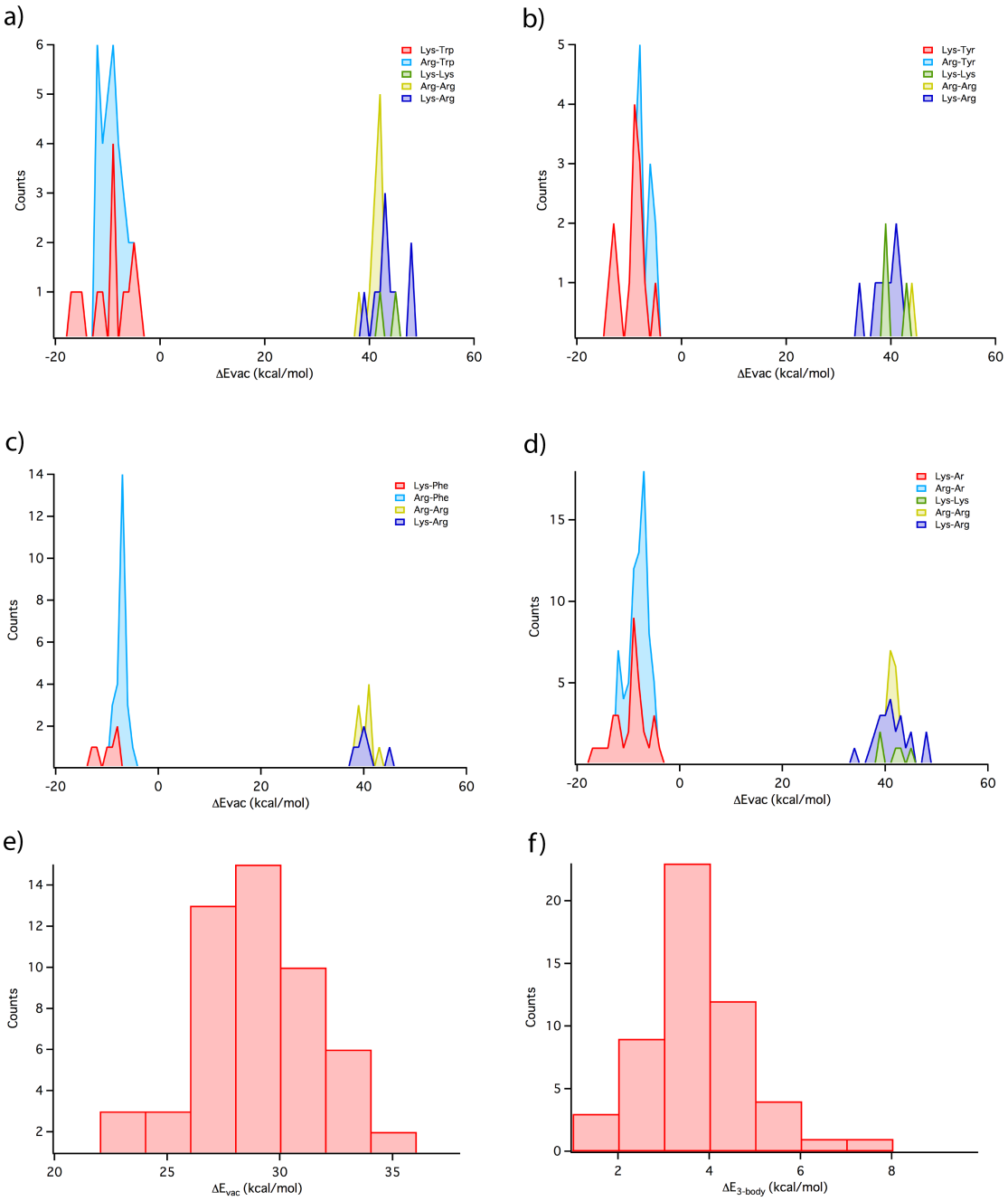

Figure 10. Interaction energies computed for cation–π–cation complexes in vacuum. a) Pair interactions in Trp complexes, b) pair interactions in Tyr complexes, c) pair interactions in Phe complexes, d) all pair interactions, e) trimer interactions and f) three-body non-additive contributions.

Because cation–cation repulsion dominates the interaction energy of the cation–π–cation complexes, the overall trimer interaction energies shown in Figure 10 are strongly positive in vacuum, ranging from 21 to 35 kcal/mol. Such energies, however, do not match the sum of pair interaction energies. In other words, nonadditivity plays an

important role in these interactions, as the three-body contribution amounts typically to 3-5 kcal/mol, and, in some cases, up to 7 kcal/mol. We examined the impact of higher-order electron correlation effects in the three-body contribution by performing CCSD/aug-cc-pVDZ calculations on the 3 complexes identified in the human growth hormone receptor (see Fig. 2). Our results pointed, however, to negligible corrections less than 0.1 kcal/mol to the SCS-MP2 estimates, as reported in Table S10 of the Supporting Information.

The finding of significant nonadditivity has important consequences for the description of cation–π–cation motifs in widely used pairwise additive, notably multipurpose macromolecular force fields for molecular simulations of proteins, as nonaddivitity is neglected in such a formalism. For instance, in order to improve the description of cation–π interactions in pairwise, additive force fields, Minoux and Chipot introduced a short-range 4-12 potential in the context of the Amber force field able to reproduce induction phenomena in an average sense.[7] Recently, Khan and co-workers showed that the Drude polarizable force field improves the description of tyrosine–choline cation–π energetics compared to the CHARMM additive force field, and proposed a modification of the Lennard-Jones terms for the latter in order to recover the missing induction effects.[65] Although such ad hoc modifications provide efficient and simple corrections, the implicit treatment of polarization effects in such formalisms becomes problematic in a cation–π–cation motif, given that the approach of the two cations on each side of the aromatic ring will partially zero out the induction contribution, whereas the 4-12 potential or the modified Lennard-Jones one would predict similar interaction energies as those found in the individual cation–π complexes. This fact reflects the neglect of non-additive contributions, like those estimated here for cation–π–cation motifs, in a pairwise additive force field.

A natural alternative to improve the description of cation–π–cation interactions would be to use an explicit polarizable force field.[24,25] In this case, common formalisms in this context, like those based on Drude oscillators or induce dipoles moments, would likely provide good descriptions of cation–π interactions, as shown in the recent study by Khan and co-workers.[65] When two cations approach the aromatic ring like in a cation–π–cation motif, however, most likely the Drude oscillators will reorient in the plane of the π–electron cloud (or the induced dipoles will partially vanish), so their

ability to describe such interactions is not warranted. A rigorous model to account for such multibody effects would probably need the inclusion of charge flows and dipolar polarizabilities, but the development of such models for solvated macromolecules is rather complex. Nevertheless, it would be interesting to explore the ability of nonadditive polarizable force fields, for example Drude[66] or AMOEBA,[67] to describe the energetics of cation–π–cation interactions, and whether they are able to account for such nonadditivity.

In a protein, the local environment can significantly impact these energies by significantly screening the electrostatic forces at play.[68] In order to investigate the impact of the environment, we adopted two strategies. First, we evaluated the impact that a polar or an apolar environment can have on cation–π–cation interactions by computing the solvation free energies in water and *n*-octanol for the individual monomers, and the trimers, based on continuum solvent MST-IEFPCM calculations[53–55] performed at the B3LYP/6-31G(d) level of theory. This allows us to estimate the effective interaction energy of a given cation–π–cation complex in solution, which is estimated as the addition of the gas phase interaction energy and the change in the solvation free energies of the complex and the separated monomers (Eq. 4). Next, we estimated the contribution of the protein environment by computing the free-energy change associated to transferring the trimer from water to the protein scaffold using the classical MM-GBSA method[56,57] based on the Generalized Born continuum solvation model (Eq. 5).

$$\Delta E_X = \Delta E_{vac} + \Delta\Delta G^X_{solv} \qquad (X: water, n-octanol) \tag{4}$$

$$\Delta E_{prot} = \Delta E_{wat} + \Delta G_{wat \rightarrow prot} \tag{5}$$

The corresponding interaction energies in *n*-octanol, water and the protein, obtained combining the vacuum SCS-MP2/CBS interaction energies with the solvation energies computed using the QM MST solvation model and the transfer free energies computed using MM-GBSA are shown in Figure 11 (see also Table S9 of the Supporting Information). As expected, the environment has a strong impact on the interaction energies, stabilizing them by 14-28 kcal/mol in *n*-octanol ($\Delta\Delta G^{oct}_{solv}$), and by 15-31 kcal/mol in water ($\Delta\Delta G^{wat}_{solv}$). In *n*-octanol, the resulting interaction energies ΔE_{oct} are still predicted to be positive in all cases, whereas in water solvation leads to

attractive interactions (ΔE_{wat}) in several complexes, even though only by up to -1.3 kcal/mol. Such calculations, nevertheless, ignore the local environment of the cation–π–cation interaction, where the presence of anionic amino acids could substantially further stabilize it. Indeed, the water to protein transfer free energies computed for the complexes in their native environment, $\Delta G_{wat \rightarrow prot}$, show an important impact of the protein environment on the interaction energies, with values ranging from -45 to 30 kcal/mol. In 40 out of the 53 interactions considered, however, the protein environment stabilizes the interaction, whereas in 13 complexes its contribution is positive.

By adding the transfer free energy to the interaction energies in water, we, thus, provide an estimate of the interaction energies for the cation–π–cation motifs in their context, ΔE_{prot}. The resulting values range from -43 to 33 kcal/mol, 34 out of the 53 complexes resulting in negative interaction energies. This trend is similar for the complexes involving Trp, Tyr and Phe, wherein the fraction of interactions that are predicted to be attractive is 61%, 64% and 69%. As noted previously, the presence of nearby anionic amino acids can considerably stabilize the cation–π–cation motifs by establishing favorable electrostatic interactions with Lys and Arg residues, and thus attenuate the repulsion between the cations. The impact of nearby anions in cation–π interactions has been the focus of several studies.[69–75] Moreover, the term salt-bridge–π interaction has been proposed to designate the contact between an aromatic ring and a planar salt-bridge, as opposed to cation–π and anion–π contacts where the cation or the anion are located above/below the aromatic ring.[75] In the set of 53 cation–π–cation motifs analyzed here, there are often one or more anionic amino acids near the complex. However, in the few cases where the anion–ring separation is comparable to the cation–ring one, there is a large displacement of the anion from the normal axis to the aromatic ring. This is probably caused by the geometric criteria used to select the cation–π–cation motifs, which favors the location of the cations over the ring.

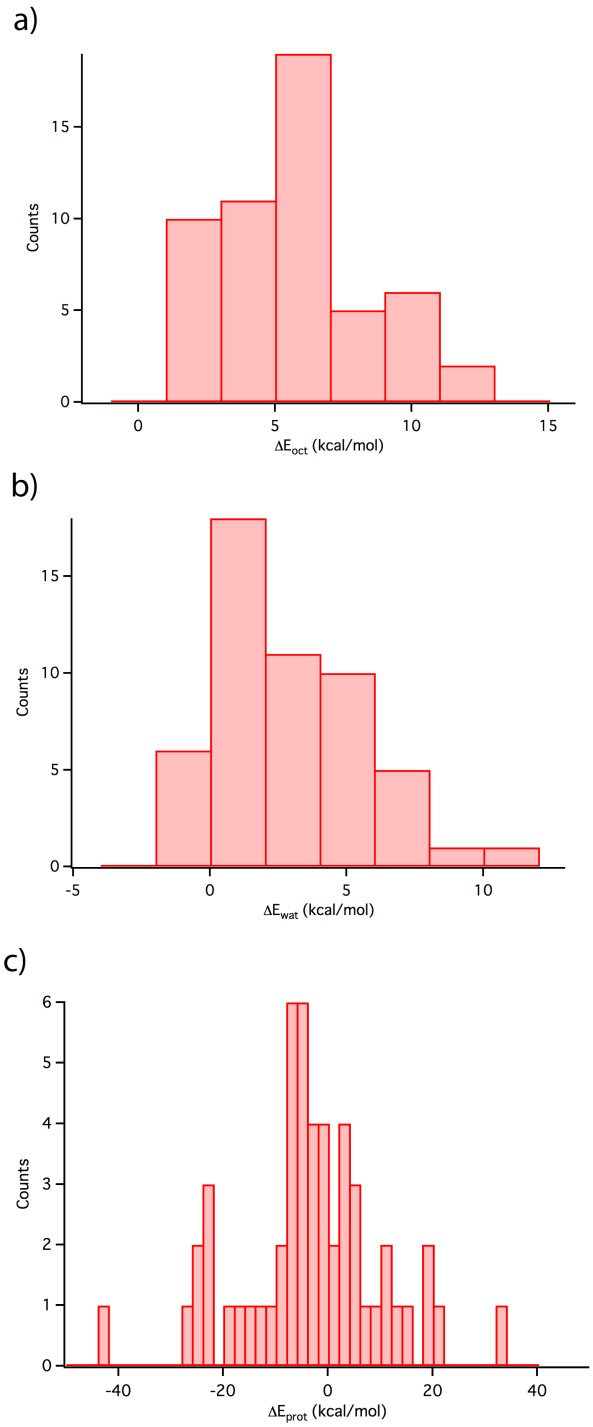

Figure 11. Effective interaction energies computed for cation–π–cation complexes in different environments. a) Apolar (*n*-octanol), b) polar (water) and c) protein environments.

The impact of negatively charged amino acids is exemplified by the network of interactions established between Glu and Asp residues and the cation–π–cation motifs in the structures of hydroxynitrile lyase and deblocking aminopeptidase, as shown in Figure 12, which translate into ΔE_{prot} values of -25.0 and -18.3 kcal/mol for these particular interactions. These results suggest that the local environment may confer to cation–π–cation interactions a stabilizing role in the tertiary structure of proteins. This possibility is supported by the high degree of conservation found for these interactions, in particular for those involving Trp and Phe. We have not found direct correlations among the average conservation score of each interaction and their interaction energies, but other functional roles could explain this finding.

Beyond individual interactions, networks of solvent-exposed cation–π interactions formed by two consecutive Arg–Trp–Arg motifs have recently been demonstrated to stabilize a miniature protein tertiary structure.[43] Here, we have explored an extended cation–π network, also exposed to the solvent, identified in the human growth hormone receptor (PDB entry 1A22), where three consecutive cation–π–cation interactions are established: Lys–Trp–Arg, Arg–Phe–Arg and Arg–Tyr–Lys (see Figure 2). We estimate the interaction energies to be equal to -16.7, -1.0 and -7.8 kcal/mol for these complexes, supporting the notion that diverse cation–π networks can be engineered in order to stabilize a tertiary structure. Future studies are required to discern the potential implications of these motifs on the intrinsic conformational flexibility and functional role of the protein.

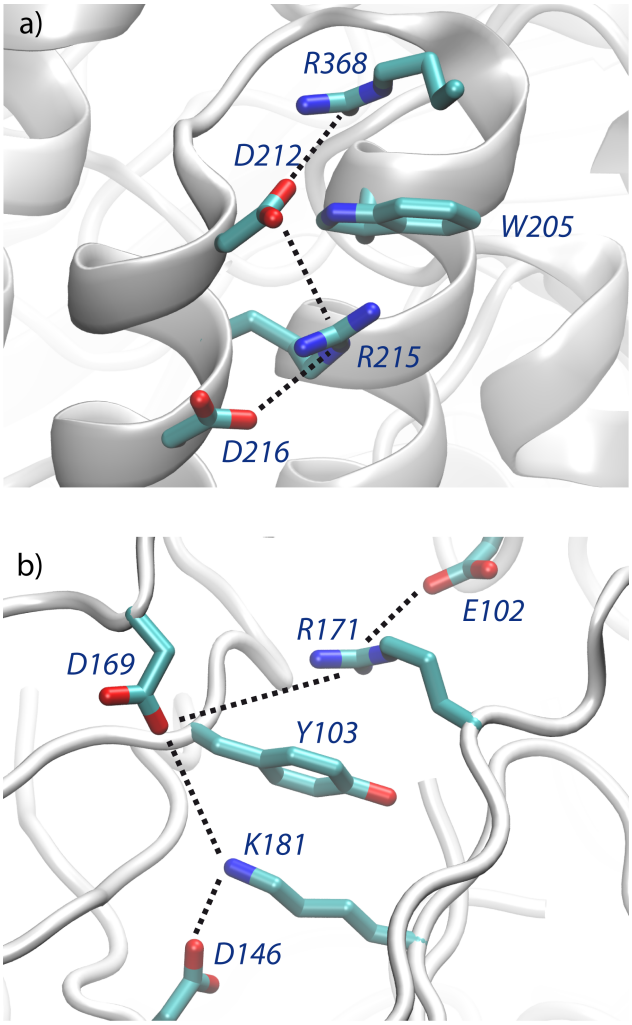

Figure 12. Illustration of the stabilizing interactions established between cation–π–cation motifs and negatively charged (Asp and Glu) amino acids in a) hydroxynitrile lyase (code 1GXS-2) and b) deblocking aminopeptidase (code 2GRE-1). More details on each interaction code can be found in Table S4-S5 of the Supporting Information.

CONCLUSIONS

We have presented a statistical analysis of the occurrence, the composition and the geometrical preferences of cation–π–cation interactions identified in a non-redundant set of protein structures taken from the PDB. Our analysis indicates that this structural motif is common in proteins, 7% of them containing at least one cation–π–cation interaction. The composition of the interactions identified, when weighted by the natural abundance of amino acids, suggests a preference to establish cation–π–cation motifs

with Trp, followed by Tyr and Phe, with approximate ratios 3:2:1, whereas no particular preference among Lys and Arg is observed regarding the cations. We have also found that cation–π–cation interactions tend to be highly conserved, hence suggesting a relevant structural or functional role. In this case, interactions involving Trp, and to a lesser extent Phe, seem to be more strongly conserved compared to complexes involving Tyr.

We have also performed an energetic analysis of a representative subset of cation–π–cation complexes combining quantum chemical SCS-MP2/CBS calculations and MST and MM-GBSA continuum solvation models. Our results point out that, whereas the interaction energy in vacuum is strongly positive, an apolar (n-octanol) or polar (water) environment can strongly screen the cation–cation repulsion, although in most cases the interaction is still predicted to be repulsive. The impact of the particular protein environment is, however, predicted to be stronger, and leads to an attractive interaction in 64% of the complexes analyzed. This result, together with the high degree of conservation of the amino acids involved in the interactions, suggests a potential stabilizing role in protein tertiary structures, as demonstrated recently for a miniature protein.[43] Finally, we found a significant degree of cooperativity among the two cation–π interactions at play. From a computational point of view, the significant contribution of such non-additive, three-body terms, which can amount up to 7 kcal/mol, challenges the suitability of standard additive biomolecular force fields for describing cation–π–cation motifs in molecular simulations of proteins.

ACKNOWLEDGEMENTS

We are grateful to the Consorci de Serveis Universitaris de Catalunya for providing access to computational resources. Financial support from the Spanish Ministerio de Economía y Competitividad (MINECO; grants CTQ2012-36195, SAF2014-57094-R and RYC2011-08918), and the Agència de Gestió d'Ajuts Universitaris i de Recerca from Generalitat de Catalunya (GENCAT; SGR2014-1189) is acknowledged. C. C. is a Serra Húnter Fellow. S. P. is a fellow of the Ciências Sem Fronteiras program of the Conselho Nacional de Desenvolvimento Ciêntifico e Tecnológico (CNPq) de Brasil (246791/2012-8). F. J. L. acknowledges the financial support by ICREA Academia.

REFERENCES

1 E. Campbell, M. Kaltenbach, G. J. Correy, P. D. Carr, B. T. Porebski, E. K. Livingstone, L. Afriat-Jurnou, A. M. Buckle, M. Weik, F. Hollfelder, N. Tokuriki and C. J. Jackson, *Nat. Chem. Biol.*, 2016, **12**, 944–950.

2 G. Kiss, N. Çelebi-Ölçüm, R. Moretti, D. Baker and K. N. Houk, *Angew. Chemie - Int. Ed.*, 2013, **52**, 5700–5725.

3 H. Kries, R. Blomberg and D. Hilvert, *Curr. Opin. Chem. Biol.*, 2013, **17**, 221–228.

4 D. A. Dougherty, *Science*, 1996, **271**, 163–168.

5 J. C. Ma and D. A. Dougherty, *Chem. Rev.*, 1997, **97**, 1303–1324.

6 E. Cubero, F. J. Luque and M. Orozco, *Proc. Natl. Acad. Sci.*, 1998, **95**, 5976–5980.

7 H. Minoux and C. Chipot, *J. Am. Chem. Soc.*, 1999, **121**, 10366–10372.

8 K. E. Riley, M. Pitoňák, P. Jurečka and P. Hobza, *Chem. Rev.*, 2010, **110**, 5023–5063.

9 I. Alkorta, F. Blanco, P. M. Deyà, J. Elguero, C. Estarellas, A. Frontera and D. Quiñonero, *Theor. Chem. Acc.*, 2010, **126**, 1–14.

10 A. Frontera, D. Quiñonero and P. M. Deyà, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 440–459.

11 D. A. Dougherty, *Acc. Chem. Res.*, 2013, **46**, 885–893.

12 A. S. Mahadevi and G. N. Sastry, *Chem. Rev.*, 2013, **113**, 2100–2138.

13 C. D. Sherrill, *Acc. Chem. Res.*, 2013, **46**, 1020–1028.

14 J. J. Fiol, M. Barceló-Oliver, A. Tasada, A. Frontera, À. Terrón and Á. García-Raso, *Coord. Chem. Rev.*, 2013, **257**, 2705–2715.

15 T. Clark, P. Politzer and J. S. Murray, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2015, **5**, 169–177.

16 A. Bauzá, T. J. Mooibroek and A. Frontera, *ChemPhysChem*, 2015, **16**, 2496–2517.

17 M. H. Kolář and P. Hobza, *Chem. Rev.*, 2016, **116**, 5155–5187.

18 A. S. Mahadevi and G. N. Sastry, *Chem. Rev.*, 2016, **116**, 2775–2825.

19 J. M. Andrić, M. Z. Misini-Ignjatović, J. S. Murray, P. Politzer and S. D. Zarić, *ChemPhysChem*, 2016, **17**, 2035–2042.

20 A. A. Rodríguez-Sanz, E. M. Cabaleiro-Lago and J. Rodríguez-Otero, *Phys.*

Chem. Chem. Phys., 2014, **16**, 22499–22512.

21 C. Rapp, E. Goldberger, N. Tishbi and R. Kirshenbaum, *Proteins Struct. Funct. Bioinforma.*, 2014, **82**, 1494–1502.

22 D. Vijay and G. N. Sastry, *Phys. Chem. Chem. Phys.*, 2008, **10**, 582–590.

23 B. U. Emenike, S. N. Bey, R. A. Spinelle, J. T. Jones, B. Yoo and M. Zeller, *Phys. Chem. Chem. Phys.*, 2016, **18**, 30940–30945.

24 P. Cieplak, F.-Y. Y. Dupradeau, Y. Duan and J. M. Wang, *J. Physics-Condensed Matter*, 2009, **21**, 333102.

25 F. J. Luque, F. Dehez, C. Chipot and M. Orozco, *Wiley Interdiscip. Rev. Comput. Mol. Sci.*, 2011, **1**, 844–854.

26 I. Soteras, C. Curutchet, A. Bidon-Chanal, F. Dehez, J. G. Ángyán, M. Orozco, C. Chipot and F. J. Luque, *J. Chem. Theory Comput.*, 2007, **3**, 1901–1913.

27 F. Dehez, J. G. Ángyán, I. S. Gutiérrez, F. J. Luque, K. Schulten and C. Chipot, *J. Chem. Theory Comput.*, 2007, **3**, 1914–1926.

28 I. Soteras, M. Orozco and F. J. Luque, *Phys. Chem. Chem. Phys.*, 2008, **10**, 2616–2624.

29 F. Archambault, C. Chipot, I. Soteras, F. J. Luque, K. Schulten and F. Dehez, *J. Chem. Theory Comput.*, 2009, **5**, 3022–3031.

30 K. Ansorg, M. Tafipolsky and B. Engels, *J. Phys. Chem. B*, 2013, **117**, 10093–10102.

31 J. W. Caldwell and P. a. Kollman, *J. Am. Chem. Soc.*, 1995, **117**, 4177–4178.

32 H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov and P. E. Bourne, *Nucleic Acids Res.*, 2000, **28**, 235–42.

33 C. Biot, R. Wintjens and M. Rooman, *J. Am. Chem. Soc.*, 2004, **126**, 6220–6221.

34 D. Quiñonero, A. Frontera, C. Garau, P. Ballester, A. Costa and P. M. Deyà, *ChemPhysChem*, 2006, **7**, 2487–2491.

35 A. S. Reddy, D. Vijay, G. M. Sastry and G. N. Sastry, *J. Phys. Chem. B*, 2006, **110**, 2479–2481.

36 A. Frontera, D. Quiñonero, A. Costa, P. Ballester and P. M. Deyà, *New J. Chem.*, 2007, **31**, 556–560.

37 Q. Li, W. Li, J. Cheng, B. Gong and J. Sun, *J. Mol. Struct. THEOCHEM*, 2008, **867**, 107–110.

38 D. Escudero, A. Frontera, D. Quiñonero and P. M. Deyà, *Chem. Phys. Lett.*, 2008, **456**, 257–261.

39 C. Estarellas, A. Frontera, D. Quiñonero and P. M. Deyà, *Chem. Phys. Lett.*, 2009, **479**, 316–320.

40 D. Vijay and G. N. Sastry, *Chem. Phys. Lett.*, 2010, **485**, 235–242.

41 R. Li, Q. Li, J. Cheng, Z. Liu and W. Li, *ChemPhysChem*, 2011, **12**, 2289–2295.

42 A. Campo-Cacharrón, E. M. Cabaleiro-Lago, J. a. Carrazana-García and J. Rodríguez-Otero, *J. Comput. Chem.*, 2014, **35**, 1290–1301.

43 T. W. Craven, M.-K. Cho, N. J. Traaseth, R. Bonneau and K. Kirshenbaum, *J. Am. Chem. Soc.*, 2016, **138**, 1543–1550.

44 T. Meyer, M. D'Abramo, A. Hospital, M. Rueda, C. Ferrer-Costa, A. Pérez, O. Carrillo, J. Camps, C. Fenollosa, D. Repchevsky, J. L. Gelpí and M. Orozco, *Structure*, 2010, **18**, 1399–1409.

45 M. H. M. Olsson, C. R. Søndergaard, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 525–537.

46 C. R. Søndergaard, M. H. M. Olsson, M. Rostkowski and J. H. Jensen, *J. Chem. Theory Comput.*, 2011, **7**, 2284–2295.

47 M. Landau, I. Mayrose, Y. Rosenberg, F. Glaser, E. Martz, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2005, **33**, W299–W302.

48 H. Ashkenazy, S. Abadi, E. Martz, O. Chay, I. Mayrose, T. Pupko and N. Ben-Tal, *Nucleic Acids Res.*, 2016, DOI: 10.1093/nar/gkw408.

49 M. J. Frisch, G. W. Trucks, H. B. Schlegel, G. E. Scuseria, M. A. Robb, J. R. Cheeseman, G. Scalmani, V. Barone, B. Mennucci, G. A. Petersson, H. Nakatsuji, M. Caricato, X. Li, H. P. Hratchian, A. F. Izmaylov, J. Bloino, G. Zheng, J. L. Sonnenberg, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, T. Vreven, J. Montgomery J. A., J. E. Peralta, F. Ogliaro, M. Bearpark, J. J. Heyd, E. Brothers, K. N. Kudin, V. N. Staroverov, R. Kobayashi, J. Normand, K. Raghavachari, A. Rendell, J. C. Burant, S. S. Iyengar, J. Tomasi, M. Cossi, N. Rega, N. J. Millam, M. Klene, J. E. Knox, J. B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R. E. Stratmann, O. Yazyev, A. J. Austin, R. Cammi, C. Pomelli, J. W. Ochterski, R. L. Martin, K. Morokuma, V. G. Zakrzewski, G. A. Voth, P. Salvador, J. J. Dannenberg, S. Dapprich, A. D. Daniels, Ö. Farkas, J. B. Foresman, J. V Ortiz, J. Cioslowski and D. J. Fox, Gaussian 09, Revision A.2. Gaussian, Inc.: Wallingford CT 2009.

50 S. Grimme, *J. Chem. Phys.*, 2003, **118**, 9095–9102.

51 D. G. Truhlar, *Chem. Phys. Lett.*, 1998, **294**, 45–48.

52 S. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.

53 C. Curutchet, M. Orozco and F. J. Luque, *J. Comput. Chem.*, 2001, **22**, 1180–1193.

54 C. Curutchet, A. Bidon-Chanal, I. Soteras, M. Orozco and F. J. Luque, *J. Phys. Chem. B*, 2005, **109**, 3565–3574.

55 I. Soteras, C. Curutchet, A. Bidon-Chanal, M. Orozco and F. J. Luque, *J. Mol. Struct.-Theochem*, 2005, **727**, 29–40.

56 D. A. Case, T. A. Darden, I. T.E. Cheatham, C. L. Simmerling, J. Wang, R. E. Duke, R. Luo, R. C. Walker, W. Zhang, K. M. Merz, B. Roberts, S. Hayik, A. Roitberg, G. Seabra, J. Swails, A. W. Götz, I. Kolossváry, K. F. Wong, F. Paesani, J. Vanicek, R. M. Wolf, J. Liu, X. Wu, S. R. Brozell, T. Steinbrecher, H. Gohlke, Q. Cai, X. Ye, J. Wang, M.-J. Hsieh, G. Cui, D. R. Roe, D. H. Mathews, M. G. Seetin, R. Salomon-Ferrer, C. Sagui, V. Babin, T. Luchko, S. Gusarov, A. Kovalenko and P. A. Kollman, AMBER 12. University of California: San Francisco, 2012.

57 A. Onufriev, D. Bashford and D. A. Case, *Proteins Struct. Funct. Bioinforma.*, 2004, **55**, 383–394.

58 V. Hornak, R. Abel, A. Okur, B. Strockbine, A. Roitberg and C. Simmerling, *Proteins-Structure Funct. Bioinforma.*, 2006, **65**, 712–725.

59 C. I. Bayly, P. Cieplak, W. Cornell and P. A. Kollman, *J. Phys. Chem.*, 1993, **97**, 10269–10280.

60 P. E. Mason, G. W. Neilson, C. E. Dempsey, A. C. Barnes and J. M. Cruickshank, *Proc. Natl. Acad. Sci.*, 2003, **100**, 4557–4561.

61 A. Grossfield, P. Ren and J. W. Ponder, *J. Am. Chem. Soc.*, 2003, **125**, 15671–15682.

62 D. Jiao, C. King, A. Grossfield, T. A. Darden and P. Ren, *J. Phys. Chem. B*, 2006, **110**, 18553–18559.

63 J. Liu, C. P. Kelly, A. C. Goren, A. V. Marenich, C. J. Cramer, D. G. Truhlar and C.-G. Zhan, *J. Chem. Theory Comput.*, 2010, **6**, 1109–1117.

64 G. Trinquier and Y. H. Sanejouand, *Protein Eng. Des. Sel.*, 1998, **11**, 153–169.

65 H. M. Khan, C. Grauffel, R. Broer, A. D. MacKerell, R. W. A. Havenith and N. Reuter, *J. Chem. Theory Comput.*, 2016, **12**, 5585–5595.

66 J. A. Lemkul, J. Huang, B. Roux and A. D. MacKerell, *Chem. Rev.*, 2016, **116**,

4983–5013.

67 Y. Shi, Z. Xia, J. Zhang, R. Best, C. Wu, J. W. Ponder and P. Ren, *J. Chem. Theory Comput.*, 2013, **9**, 4046–4063.

68 A. Warshel, P. K. Sharma, M. Kato and W. W. Parson, *Biochim. Biophys. Acta, Proteins Proteomics*, 2006, **1764**, 1647–1676.

69 Y. Inoue, S. Sugio, J. Andzelm and N. Nakamura, *J. Phys. Chem. A*, 1998, **102**, 646–648.

70 P.-O. Norrby and T. Liljefors, *J. Am. Chem. Soc.*, 1999, **121**, 2303–2306.

71 S. D. Zarić, D. M. Popović and E.-W. Knapp, *Chemistry (Easton).*, 2000, **6**, 3935–3942.

72 S. E. Thompson and D. B. Smithrud, *J. Am. Chem. Soc.*, 2002, **124**, 442–449.

73 V. Dvornikovs and D. B. Smithrud, *J. Org. Chem.*, 2002, **67**, 2160–2167.

74 M. A. Anderson, B. Ogbay, R. Arimoto, W. Sha, O. G. Kisselev, D. P. Cistola and G. R. Marshall, *J. Am. Chem. Soc.*, 2006, **128**, 7531–7541.

75 M. Mitra, P. Manna, S. K. Seth, A. Das, J. Meredith, M. Helliwell, A. Bauzá, S. R. Choudhury, A. Frontera and S. Mukhopadhyay, *CrystEngComm*, 2013, **15**, 686–696.